

A Scalable Approach to Data Management Education of Graduate Students

UNIVERSITY OF MINNESOTA
LIBRARIES

Jon Jeffryes (jeffryes@umn.edu) and Lisa Johnston (ljohnsto@umn.edu)
University of Minnesota – Twin Cities



Introduction

As part of the Data Information Literacy (DIL) project, funded by a grant from the Institute of Museum and Library Services (IMLS), the University of Minnesota (UMN) Libraries partnered with Purdue University, Cornell University, and the University of Oregon to identify:

- ✓ What data management and curation skills are needed by graduate students in the sciences to take advantage of opportunities in e-science?
- ✓ How can academic librarians contribute their expertise in information organization, dissemination and preservation to better serve modern science?

The UMN team worked with a structures research group in the Civil Engineering Department.

Methodology

Part 1. In-Depth Interviews (Spring 2012)



- ✓ 90-120 minutes
- ✓ 4 graduate students
- ✓ 1 faculty member
- ✓ Interview tools: z.umn.edu/dil

Based on our findings¹ and an extensive literature review and environmental scan, we developed online course in Data Management.

Part 2. Online Course (Fall/Sp 2012-13)



- ✓ ScreenFlow and YouTube
- ✓ Google Docs, Forms, and Sites

Our e-learning approach² to teaching data management skills can be scaled to a variety of disciplines - made easier with tools such as Google Sites allowing us to reuse the content.

Case Study: Structural Engineering Graduate Student Needs

DIL Need	Backup and Security	Documentation	Ownership	Sharing	Long-term Access
Evidence	Students stored data on laptops that do not always include backup strategies.	Students graduate and passed their data to the next student without context.	State-funded research projects created ownership confusion.	Students did not recognize the reuse potential of their data.	Some data was collected using proprietary file-formats.

Anatomy of the Online Data Management Course

Learning objectives were identified from graduate student needs assessment.

7 modules map to Data Management Plan requirements of NSF and NIH.

Resources point to services outside the library.

Lesson video is self-contained and can be viewed on multiple devices.

Example Module from UMN Data Management Course

2. Data to be Managed

The data that we produce in our research can grow dramatically. Without a clear plan on how to manage our data before we begin the collection process, our files, raw data points, documentation logs, and simulations can quickly overwhelm and become a "data deluge." At the end of this module you will be able to:

- define what information you will be managing, and
- describe the importance of appropriate storage and backup solutions to securely house these data.

Lesson 2: Data Types, Storage and Backup

Instructions: Watch this 14 mini video tutorial, and then complete complete assignment 2.

Assignment 2

In your DMP (check this link), complete section 2 "Data Types" by creating a data inventory for your research project (data, project files, documentation, etc.) to describe what data you will manage. Also detail your data storage and backup plan to avoid potential loss of data.

Other questions to consider include:

- How will data be collected? In what file formats?
- How to document issues/problems in the data collection process?
- Will it be reproducible? What would happen if it got lost or became unusable later?
- How much data, and at what growth rate? How often will it change?
- Are there tools or software needed to create/process/visualize the data?
- Will you use pre-existing data? From where?

Data Management Plan (DMP) TEMPLATE

Data Management Plan

Name of student/researcher: _____ Your Name: _____

Name of group/project: _____ Project Name or Research Lab (for group plans): _____

Primary investigator: _____

Project duration: _____ Start: MM-DD-YYYY End: MM-DD-YYYY

Class number: _____

1. Introduction

The research project described in this data management plan (DMP) ...

Assessment is formative and students asynchronously complete a DMP template that instructors can view in a shared Google Doc.

Enrollment in the First Year of the Data Management Course

Online Course Enrollment and Completion Rates (Fall/Spring 2012-13)				
	Enrolled	Discipline	Completion Rate	Notes
Fall 2012*	11	Structural Engineering	45% (5)	Two completed on time and three students deferred to the Spring.
Spring 2013**	47	14 Sci/Eng Departments	13% (6)	Optional in-person lecture brought 13 participants to the library.
Total	58	STEM Disciplines	19% (11)	Structural Engineering (6) Chemistry (1) Entomology (2) Applied Economics (1) Forestry Resources (1)

Participants that completed the course received a certificate and the Data Management Course logged into their UMN Training History.

Fall 2012

Open to Structures graduate students (~20 total) in the Civil Eng Dept.

Spring 2013

We broaden the course content to offer to all science & engineering graduate students.

Results

The open online-course format provided a platform to launch a scalable data information literacy curriculum across the STEM disciplines.

- ✓ 58 STEM graduates students enrolled
- ✓ 14 academic departments represented
- ✓ 11 students completed a written DMP

An evaluation survey of graduates (64% response) demonstrated a high-level of satisfaction in the course"

"This course gave me good techniques which I will not only be able to implement in my current research in addition to what I have already been doing, but also use them in the rest of my career."

Six-month Follow-Up Assessment

Two Fall graduates have responded (100% response) to a follow-up survey to help determine how the DIL skills were applied.

- ✓ 2/2 created a file-naming structure
- ✓ 1/2 employed a metadata schema
- ✓ All modules of the Data Management Course were found to be either "Useful" or "Very Useful"

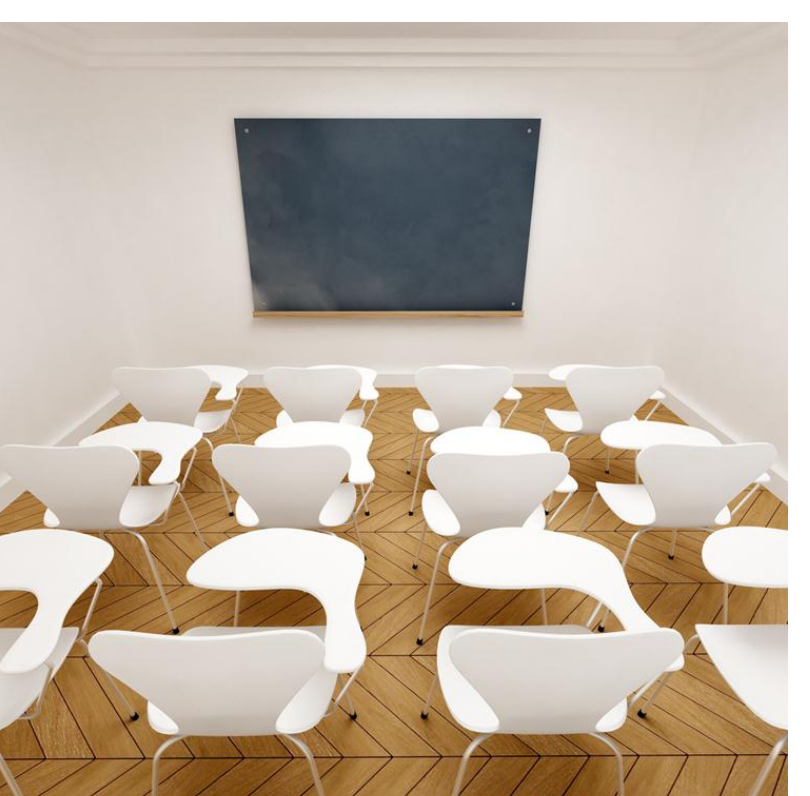


Next Steps

"I really liked the in-person lecture. Made it easy to set aside one block of time to go through all the information and have staff on-hand to answer questions."

Fall 2013 Workshop Series: "Flipped Classroom" Approach

- ✓ Five sessions focus on Module topics 2-6
- ✓ Ready to use online content
- ✓ Hands-on training with actual data sets
- ✓ Questions and feedback opportunities



Want more? <http://z.umn.edu/datamgmt>



- Johnston, L. and Jeffryes, J. (2013). "Data Management Skills Needed by Structural Engineering Students" J. Prof. Issues Eng. Educ. Pract., 10.1061/(ASCE)EI.1943-5541.0000154 (Feb. 13, 2013). [http://ascelibrary.org/doi/abs/10.1061/\(ASCE\)EI.1943-5541.0000154](http://ascelibrary.org/doi/abs/10.1061/(ASCE)EI.1943-5541.0000154)
- Jeffryes, J. and Johnston, L. (2013). "An E-Learning Approach to Data Information Literacy Education." 2013 ASEE Annual Conference (Atlanta). <http://www.asee.org/public/conferences/20/papers/6956/view>